Comment

Sleeping dogs

Gregory A Petsko

Address: Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA 02454-9110, USA.
E-mail: petsko@brandeis.edu

Published: 23 October 2003

Genome ***Biology*** 2003, **4:**120

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2003/4/11/120

'Let sleeping dogs lie' is a common English-language expression. It means, of course, to let well enough alone; that it's advisable not to disturb someone - or something - that might bite you if you do. Like many old aphorisms, there is much wisdom in it, and there have been a number of times when I wish I had followed it.

As I write this, two dogs are sleeping on the sofa next to me. I won't wake them, not because they'd bite - they're the gentlest dogs in the world and wouldn't bite anyone - but because they look so cute and peaceful snuggled together. They have no idea that last week the juggernaut of comparative genomics rolled round to them. On 26 September 2003, a joint team from The Institute for Genome Research and The Center for the Advancement of Genomics, both in Rockville, USA, and headed by Claire Fraser and Craig Venter, respectively, announced a 1.5X whole-genome sequence of the domestic dog (specifically Shadow, Claire and Craig's pet male standard poodle; see Kirkness *et al.*, *Science* 2003, **301:**1898-1903).

The dog genome sequence represents a landmark in the genomics era for several reasons. As the sequence makes clear, the dog is the closest relative to man yet to have a mostly complete draft genome sequence determined. The attempt to do it on the cheap, with minimal coverage, turned out to be surprisingly successful, presaging a flood of quick-and-dirty mammalian genome sequences in the near future. Another reason is the enormous, well-cataloged phenotypic variation of the canine: in the more than 100 centuries since the first canids were domesticated, dogs have been bred to display, in over 400 well-defined genetic sub-types ('breeds'), a huge range of morphological and behavioral characteristics that can now, in principle, be linked to their genes.

The strategy used to obtain the 1.5X whole-genome dog sequence is likely to become a model for future draft sequencing efforts. It yielded contiguous sequences (contigs) too small to extend across chromosome-length distances without a physical or genetic map; happily, there was already a radiation hybrid map that could be used to anchor the sequences to their positions in the 40 dog chromosomes. Most of the coding sequences were fragmentary, but with the aid of the human and mouse genomes it was possible to determine that about 80% of human genes have identifiable homologs in the dog. As the database of complete, high-coverage mammalian genome sequences grows (the public genome project should have a 6.5X dog genome sequence in the future, and similar efforts for chimp and cow are far advanced), future low-coverage sequences will have even more reference genomes to aid in assembly, alignment and interpretation. While the present dog genome sequence makes it clear that high-coverage sequencing is essential for the important organisms, it also demonstrates that useful information for comparative genomics and organismal biology can be obtained relatively cheaply. Since there are about 5,000 different known species of mammal, we can also conclude that the sequencing programs are not likely to end any time soon!

Humans have a higher content of repetitive DNA in their genomes (46%) than either mice (38%) or dogs (31%). Yet, even though only 2% of the dog genome is believed to code for proteins, more than 4% of the intergenic sequences are conserved between dog and human. Whether these conserved regions are functional remains to be shown, but clearly one reason for sequencing a number of mammalian genomes is that any functional constraints should eventually be apparent, and we might finally figure out what some of that 'junk' DNA is really for. Another interesting piece of information to emerge from the 1.5X dog sequence is that the overall mutation rate of the dog genome appears to be about the same as it is for humans; mice seem to have a mutation rate that is twice as fast. Given this difference, it is not surprising that the overall sequence similarity between the dog and human genome is higher than that between mouse and human or mouse and dog. Of the 24,567 annotated human genes, the dog has clearly detectable orthologs

for more than 18,000 (about 80%), and given the fragmentary nature of the dog genome sequence it seems certain that this number will eventually get much larger.

The recent dog genome paper presents data to support the view that the dog lineage was the first to diverge from the common carnivorous ancestor of dogs, mice and humans. Dogs, like mice, have a much larger number of olfactory receptor genes than humans, but surprisingly the mouse has the larger number, suggesting that those cute drug-sniffing dogs we see at airports should perhaps be replaced by mice on leashes.

Dogs are unlikely to become a major model organism: most of the tools of mammalian genetics are not yet available for the canine and most people, myself included, would rather see them as companions than research tools. But the dog genome sequence may nevertheless shed light on two areas of human biology: genetic diseases and behavior. Because of the huge veterinary literature about man's best friend, we know of at least 350 genetic diseases in the dog with human counterparts. Since a number of BAC sequences can already be found in the GenBank database from other breeds of dog, Fraser, Venter and colleagues were able to do a preliminary comparison with the standard poodle genes. Interestingly, they found examples of numerous sequences that differed only by the insertion of a short interspersed nuclear element (SINE). One SINE in particular, which apparently derives from a lysine tRNA sequence, represents 7% of the dog genome and has homologs in all carnivores. A single subfamily of this SINE with a consensus length of 189 bases has almost a quarter of a million copies in the dog genome. About 16,000 of these are estimated to be bimorphic, in contrast with fewer than 1,500 bimorphic SINES in the human population. If one of these mobile genetic elements becomes inserted in a gene, it can have significant consequences: the insertion of SINEs into the hypocretin/orexin-receptor-2 (Hcrtr2) gene in Labrador retrievers and other dogs causes narcolepsy, a chronic neurologic disorder characterized by excessive daytime sleepiness. (As one who lives with a Labrador retriever, I can only ask: how could they tell?)

But it is the possible value of the dog for understanding the genetic basis of behavior that has always intrigued biologists. The 400 breeds of dog display an enormous range of phenotypes, especially behavioral differences. The dog genome-sequence team speculates that this diversity may be largely due to the abundance of bimorphic mobile genetic elements. If so, it may be relatively easy to identify genes responsible for many different behaviors, and eventually to alter them at will. I can see such differences every day in the two dogs on my sofa. Mink, the 100-pound chocolate Labrador retriever, has qualities that anyone would want in a friend. He's brave, friendly, intelligent, calm and incredibly generous. He's also lazy. Clifford, the 20-pound mixed breed (half cocker spaniel, half poodle) sleeping next to him is not only physically very unlike his stepbrother but also completely different in character. He's selfish, greedy, fundamentally cowardly, not as bright, and generally rambunctious. It's tempting to believe that the world would be a better place if there were more people like Mink and fewer with the qualities of Clifford, but I don't think that's necessarily true. I love Clifford just as much as Mink, not in spite of his peccadilloes but because of them. The contrast between their two characters and temperaments is a constant source of delight. Without Clifford to prod and provoke him, Mink would be lazier and maybe even a bit boring. Without Mink to look after him and provide a contrast, Clifford would get in a lot more trouble and be less amusing.

Understanding the origins of behavior is apt to tempt some people to try to shape it to their own view of what is desirable. I'm not sure that we humans have the wisdom to do that. A world without selfishness may seem idyllic, but where does ambition end and selfishness begin? A world without ambition would be a world without accomplishments. Bravery is valuable, but is the absence of caution a good idea, and could we ever engineer one without the other? The world, I think, needs not the sameness of genetically determined 'goodness', whatever that is, but different kinds of people with contrasting characteristics, like Mink and Clifford. They provide the richness of life and are necessary for human progress. Maybe greed, selfishness, foolhardiness and other 'negative' characteristics are the price we have to pay as a species for the existence of determination, overachievement, courage, and a host of other traits we find desirable. Maybe, as some philosophers have suggested, good can't exist without evil. I don't know if these things are true, but how can we afford to take the chance? Manipulating behavior genetically seems to me the kind of thing that can wake up and bite you. I think this is one sleeping dog we would do well to let lie. Now, if you'll excuse me, it's time for them to take me on my afternoon walk.